The European Journal of International Law Vol. 32 no. 1

Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law

Barrie Sander*

Abstract

Once hailed as beacons of democracy, social media platforms such as Facebook, YouTube and Twitter now find themselves credited with its decay. Amidst a rising global techlash and growing calls for digital constitutionalism, this article develops a typology of the diverse forms of governance enabled by social media platforms and examines the contestability of human rights law in addressing the accountability deficits that characterize the platform economy. The article examines two interrelated forms of social media governance in particular: content moderation, encompassing the practices through which social media companies determine the permissibility and visibility of online content on their platforms; and data surveillance, encompassing the practices through which social media companies process personal data in accordance with their extractivist business models. Recognizing that human rights law is a vocabulary of governance with the potential to both restrain and legitimate particular relations of power within the platform economy, this article critically examines two rival conceptions of human rights law – marketized and structural – that may be relied upon to address the accountability shortfalls that pervade the contemporary social media ecosystem. The article ultimately argues in favour of a more structural conception of human

* Assistant Professor of International Justice, Faculty of Governance and Global Affairs, Leiden University, The Netherlands. Email: b.j.sander@luc.leidenuniv.nl. A draft of this article was discussed at the EJIL's 30th Anniversary Symposium: International Law and Democracy Revisited. I am grateful to the discussants Diane Desierto and Benedict Kingsbury for their constructive feedback. I would also like to thank Molly Land, evelyn douek and Paddy Leerssen for their comments on subsequent drafts of the article. I am also grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for providing funding that enabled this research to be conducted. Any errors are mine alone.

rights law, one characterized by an openness to positive state intervention to safeguard public and collective values such as media pluralism and diversity as well as a systemic lens that strives to take into account imbalances of power in the social media ecosystem and the effects of state and platform practices on the social media environment as a whole.

1 Introduction

Over the course of the past decade, control over the content layer of the online environment has become increasingly concentrated in a small number of social media companies – private enterprises that govern user-generated content on digital platforms, typically for profit.[1] During their start-up phase, it was conventional wisdom for social media platforms such as Facebook, YouTube and Twitter to be celebrated as forms of 'liberation technology' that could empower individuals to communicate, mobilize protest, scrutinize government and expose wrongdoing.[2] More recently, however, social media's honeymoon period has come to a close. Amidst a growing number of public controversies, social media companies are now facing a global 'techlash',[3] characterized by rising anxieties over the corrosive effects of their platforms on democratic processes around the world.[4] Once hailed as a boon to democracy, social media companies now find themselves under scrutiny for its decay.[5]

Democracy is a highly contested concept that has proven notoriously resistant to definition.[6] Indeed, for some, contestation about the meaning of democracy is at its very core.[7] Rather than attempting to resolve this debate, this article argues that today's social media ecosystem poses a major challenge to what Hilary Charlesworth refers to as 'the basic impulse for democracy' – namely, 'accountability for the use of power and the prevention of its arbitrary exercise'.[8] A testament to this challenge is the kinetic growth of interest in 'digital constitutionalism', a term encompassing 'a constellation of initiatives that have sought to articulate a set of political rights, governance norms, and limitations on the exercise of power on the Internet'.[9] In a climate of increasing urgency for digital constitutionalism, this article seeks to make two

1 T. Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (2018), at 18–21.

2 Diamond, 'Liberation Technology', 21 *Journal of Democracy* (2010) 69, at 70.

3 'Internet Firms Face a Global Techlash', *The Economist*, 10 August 2017.

4 Tufekci, 'How Social Media Took Us from Tahrir Square to Donald Trump', *MIT Technology Review* (14 August 2018).

5 N. Persily, *The Internet's Challenge to Democracy: Framing the Problem and Assessing Reforms* (2019), at 8.

6 See generally Daly, 'Democratic Decay: Conceptualising an Emerging Research Field', 11 *Hague Journal on the Rule of Law* (2019) 9.

7 Dryzek, 'Can There Be a Human Right to an Essentially Contested Concept? The Case of Democracy', 78 *Journal of Politics* (2016) 357, at 363.

8 Charlesworth, 'International Legal Encounters with Democracy', 8 *Global Policy* (2017) 34, at 40.

9 Redeker, Gill and Gasser, 'Towards a Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights', 80 *International Communication Gazette* (2018) 302, at 303. See also De Gregorio, 'The Rise of Digital Constitutionalism in the European Union', *International Journal of Constitutional Law* (13 April 2021).

contributions to the bourgeoning literature on the relationship between social media platforms and democracy.

The first contribution is *taxonomic* in nature. Like all forms of technology, social media platforms are sites of governance that mediate and constitute relationships of power and control between different actors.[10] In order to illuminate the accountability deficits that pervade the contemporary social media ecosystem, this article examines two interrelated forms of social media governance: *content moderation*, encompassing the practices through which social media companies determine the permissibility and visibility of online content on their platforms; and *data surveillance*, encompassing the practices through which social media companies process personal data in accordance with their extractivist business models. Drawing on existing scholarship, this article elaborates two key distinctions related to these forms of governance: first, a distinction between *liability-driven* content moderation that social media companies are incentivized to undertake in accordance with formal state legislation and *context-driven* content moderation that social media companies undertake under the influence of more informal pressures exerted by a wider range of actors;[11] and second, a distinction between platforms acting as '*surveillance intermediaries*' situated between the state and user data and '*surveillance principals*' that process user data in accordance with their own commercial interests.[12] By surfacing these different manifestations of social media governance, this article illuminates how social media platforms constitute 'sites of encounter' that establish and sustain relations of power and control between different actors.[13]

The second contribution is *doctrinal* in nature. Whereas efforts towards digital constitutionalism have often been aspirational, this article critically examines the potential and limits of existing human rights law (HRL) to address the accountability deficits associated with the structures of governance of the social media age. An important context for this discussion is the growing scholarly attention that has been devoted to understanding the relationship between the ascendancy of HRL and the parallel entrenchment of neoliberal structures of governance around the world, including processes of privatization, financialization and the protection of capital from democratic demands for social redistribution and protection.[14] A significant development within the current era of neoliberalism has been the explosive growth of the digital economy and the rise of informational capitalism.[15] Indeed, the emergence of market-dominant social media platforms as essential channels of public communication underpinned by

[10] Balkin, 'Free Speech in the Algorithmic Society', 51 *University of California Davis Law Review* (*UCDLR*) (2018) 1149, at 1157–1160.

[11] See similarly Jørgensen and Pedersen, 'Online Service Providers as Human Rights Arbiters', in M. Taddeo and L. Floridi (eds), *The Responsibilities of Online Service Providers* (2017) 179 (distinguishing between mandatory and voluntary measures).

[12] Rozenshtein, 'Surveillance Intermediaries', 70 *Stanford Law Review* (2018) 99; Cohen, 'Law for the Platform Economy', 51 *UCDLR* (2017) 133, at 191–199.

[13] Cohen, *supra* note 12, at 136.

[14] See generally Kapczynski, 'The Right to Medicines in an Age of Neoliberalism', 10 *Humanity* (2019) 79.

[15] See generally Balkin, 'The Political Economy of Freedom of Speech in the Second Gilded Age', *Law and Political Economy* (4 July 2018); and Kapczynski, 'The Law of Informational Capitalism', 129 *Yale Law Journal* (2020) 1460.

elaborate architectures of content control and data surveillance has even been characterized as one of 'neoliberalism's greatest triumph[s]'.[16] Situated in this context, this article examines the role of HRL in addressing the accountability deficits that characterize today's 'increasingly privately controlled, neoliberal communication sphere'.[17]

Conceiving of HRL as a field of contestation and struggle, this article begins from the premise recently articulated by Amy Kapczynski that HRL is 'no mere bystander in our neoliberal age' but 'inevitably entangled with neoliberal legality' and that, therefore, the struggle for the meaning of HRL 'matters for those who wish to challenge the prevailing order, not only because it could help advance real change, but also because it could forestall it'.[18] In this vein, this article argues that the open-textured and context-independent nature of human rights obligations renders them open to different ways of being understood and interpreted in the context of social media governance.[19]

On the one hand, HRL may be understood pursuant to what Upendra Baxi famously termed a 'market-friendly, human rights paradigm' that reinforces existing neoliberal structures of governance.[20] Marketized conceptions of HRL are premised on the laissez-faire free market assumption that the primary aim of HRL is to protect individual choice and agency against state intervention. Such conceptions tend to adhere to a form of abstract individualism that neglects power asymmetries between individual users and other actors that participate in the social media ecosystem and pays minimal attention to the systemic effects of state and platform practices on the social media environment as a whole. On the other hand, it is possible to envisage a more structural conception of HRL that relies on 'a structural understanding of power relations as providing a basis for legal intervention'.[21] Applied to the contemporary social media ecosystem, structural conceptions of HRL are characterized by a greater openness to positive state intervention as a means of safeguarding public and collective values such as media pluralism and diversity. In addition, structural conceptions tend to adopt more systemic perspectives that strive to take into account imbalances of power in the social media ecosystem as well as the effects of state and platform practices on the social media environment as a whole.[22]

16 Starr, 'How Neoliberal Policy Shaped the Internet – and What to Do About It Now', *The American Prospect* (2 October 2019).

17 S. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (2018), at 92.

18 Kapczynski, *supra* note 14, at 82.

19 Cotula, 'Between Hope and Critique: Human Rights, Social Justice and Re-Imagining International Law from the Bottom Up', 48 *Georgia Journal of International and Comparative Law* (2020) 473, at 478–485.

20 Baxi, 'Voices of Suffering and the Future of Human Rights', 8 *Transnational Law and Contemporary Problems* (1998) 125, 163–164.

21 Davidson, 'The Feminist Expansion of the Prohibition of Torture: Towards a Post-Liberal International Human Rights Law?', 53 *Cornell International Law Journal* (2019) 109, at 114.

22 See also douek, 'The Limits of International Law in Content Moderation', *SSRN* (11 October 2020), at 30 (suggesting the need for 'a more systemic view of rights' that takes into account the inevitability of errors of content moderation at scale); douek, 'Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability', *SSRN* (4 October 2020), at 38 (arguing that 'content moderation is a task of systemic balancing: interests are balanced and error rates are rationalized at the level of system design') (emphasis in original). For a related constitutional perspective, see generally De Gregorio, 'From Constitutional Freedoms to the Power of the Platforms: Protecting Fundamental Rights Online in the Algorithmic Society', 11 *European Journal of Legal Studies* (2019) 65; and De Gregorio, 'Democratising Online Content Moderation: A Constitutional Framework', 36 *Computer & Security Law Review* (2020).

With these differing conceptions of HRL in mind, the present article aims not only to surface the contestability of HRL in the social media governance context but also to demonstrate the importance of moving towards a more structural understanding of HRL in order to begin to close the accountability deficits associated with content moderation and data surveillance in the contemporary platform economy.

Before proceeding, however, it is important to emphasize two limitations to the present inquiry. First, this article does not attempt to examine the relationship between social media platforms and all human rights; in the interests of space, the article focuses primarily on the rights to freedom of expression and privacy, two rights that have been particularly affected by social media governance. Second, the article examines the application of HRL – encompassing binding human rights obligations of states developed at the international, regional and domestic levels – to the exclusion of the broader set of non-binding human rights norms, including the corporate responsibility to respect articulated in the United Nations' Guiding Principles on Business and Human Rights.[23] This is not to diminish the importance of the latter set of norms, the application of which in the social media context is left for detailed examination elsewhere.[24]

2 Content Moderation

The kinetic rise of social media platforms over the course of the past decade has fundamentally transformed how individuals and groups interact around the world, shifting the public sphere away from the few-to-many mass media model of communication towards a many-to-many structure in which enormous numbers of people have become both contributors and consumers of public speech.[25] Importantly, although social media companies are not responsible for producing the bulk of the content that appears on their platforms, they still make important decisions concerning both its permissibility and visibility. A central function of social media companies is content moderation, the practice of determining which categories of content are allowed and prohibited on their platforms (*content gatekeeping*) and how content is ranked and amplified (*content organizing*).[26] To moderate content, social media companies rely on a mixture of architectural design and platform rules, enforced by systems that combine community flagging, data-fuelled algorithms and human review, the precise organizational structure of which varies depending on the size, resources, purpose and culture of the platform.[27]

[23] Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework, UN Doc. HR/PUB/11/04 (2011).

[24] See generally Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation', 43 *Fordham International Law Journal* (2020) 939.

[25] E.B. Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (2015), at 15.

[26] Gillespie, *supra* note 1, at 18.

[27] See generally Caplan, 'Content or Context Moderation: Artisanal, Community-Reliant, and Industrial Approaches', *Data and Society* (2018), available at https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf; Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech', 131 *Harvard Law Review* (2018) 1598; Gillespie, *supra* note 1.

In recent years, democratic concerns with content moderation on social media platforms have arisen from at least two perspectives. On the one hand, concerns have been raised that social media companies have been intervening too aggressively in the moderation of content. Prominent examples include the removal of thousands of YouTube channels documenting human rights violations in Syria,[28] the takedown of Facebook conversations discussing online harassment[29] and the deletion of a famous Vietnam war photo from Facebook.[30] These interventions are sometimes mistakes and sometimes intentional but generally made with little transparency, due process, accountability or oversight.

At the same time, anxieties concerning the over-removal of content in certain contexts have coincided with mounting demands on social media companies to intervene more assertively in others. In particular, concerns have been raised that the moderation architecture of today's leading social media companies renders platforms ripe for 'listener-targeted speech control' by organized actors.[31] For example, the failure of social media companies to tailor their moderation systems to local contexts, including the subtleties of different languages and culturally coded forms of expression, has left their platforms vulnerable to targeted online hate and disinformation campaigns.[32] At the same time, the reliance of social media companies on algorithmic data analytic recommender systems that seek to maximize user engagement by serving users with 'relevant' content – both in the form of organic posts and paid advertisements – has resulted in the prioritization of emotionally charged forms of expression and the facilitation of clandestine forms of behavioural microtargeting.[33]

In practice, two forms of organized information campaigns have become particularly prevalent on social media platforms: first, *reverse censorship* operations where organized actors flood platforms with content designed to drown out disfavoured content or discredit particular sources of information and, second, *trolling* operations where coordinated actors attempt to provoke and/or silence users through false or inflammatory content. While these operations are not new to democratic politics, social media platforms have enabled a shift in their scale and organization. Freedom House, for example, has documented the rise of 'hyperpartisan online mobs', which 'lace their political messaging with false or inflammatory content, and coordinate its dissemination across multiple platforms'.[34] Tactics range from *amplifying organic posts* through fraudulent or automated accounts, hyper-partisan alternative news channels and

[28] See generally R. Al Jaloud *et al.*, *Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content* (2019).

[29] I. Oluo, 'Facebook's Complicity in the Silencing of Black Women', *Medium* (2 August 2017).

[30] 'Facebook Deletes Norwegian PM's Post as "Napalm Girl" Row Escalates', *The Guardian* (9 September 2016).

[31] Wu, 'The First Amendment Obsolete', in *Emerging Threats Series, Knight First Amendment Institute* (2017) 1, at 15.

[32] Sander, 'Mass Atrocities in the Age of Facebook: Towards a Human Rights-Based Approach to Platform Responsibility', *Opinio Juris*, Parts 1 and 2, (16–17 December 2019).

[33] See generally Cobbe and Singh, 'Regulating Recommending: Motivations, Considerations, and Principles' 10 *European Journal of Law and Technology* (*EJLT*) (2009).

[34] Freedom House, *Freedom on the Net 2019: The Crisis of Social Media* (2019), at 1.

paid social media personalities with sizeable followings,[35] to *online political microtargeting* whereby platform advertising services are relied upon to tailor and target messages at narrow categories of individuals for political ends.[36] These types of operations are constantly evolving, both in terms of the tactics they rely upon and the advances in technology they take advantage of, including, for example, the hyper-realistic digital falsification of images, audio or videos known as 'deep fakes'.[37]

Although the precise impact of these tactics in particular societal contexts is difficult to measure,[38] growing awareness of the different ways in which the content amplification and audience targeting systems that underpin platform business models may be leveraged has generated a range of democratic concerns.[39] For example, concerns have been raised that organized information campaigns that spread false or inflammatory content may undermine the ability of citizens to select their preferred political candidates on the basis of reliable information, drown out particular voices in public debates, deter citizens from engaging in public debates or standing for public office, discourage particular societal groups from voting, promote division and distrust amongst voters and/or undermine confidence in the integrity of a vote.[40] Significantly, these types of campaigns often seek to spread content in ways that take advantage of human cognitive and emotional biases in order to exacerbate discord or sow confusion within a particular community or society.[41] In addition, information campaigns may also undermine public trust in the veracity of online content more generally, generating a 'liar's dividend' that makes it easier for individuals and groups to cast doubt on the authenticity of online information.[42]

Reflecting on its ability to segment prospective voters into distinct groups, online political microtargeting may encourage political campaigns to pander to narrow issues with high emotional appeal at the expense of sustaining a consistent and unifying theme or vision of government for all citizens.[43] In turn, citizens may be nudged to respond to narrowly tailored issues at the expense of the larger needs of society.[44] Microtargeting may also cause political communication to become increasingly

[35] *Ibid.*, at 6–9.

[36] See generally Borgesius *et al.*, 'Online Political Microtargeting: Promises and Threats for Democracy', 14 *Utrecht Law Review* (2018) 82; S. Vaidhyanathan, *Anti-Social Media: How Facebook Disconnects Us and Undermines Democracy* (2018), ch. 6.

[37] See generally Chesney and Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', 107 *California Law Review* (*CLR*) (2019) 1753.

[38] Y. Benkler *et al.*, *Network Propaganda: Manipulation, Disinformation and Radicalization in American Politics* (2018), at 276.

[39] See generally N. Maréchal and E.R. Biddle, *It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge* (2020); N. Maréchal, R. MacKinnon and J. Dheere, *Getting to the Source of Infodemics: It's the Business Model* (2020).

[40] C. Tenove *et al.*, *Digital Threats to Democratic Elections: How Foreign Actors Use Digital Techniques to Undermine Democracy* (2018), at 26–32.

[41] Lin and Kerr, 'On Cyber-Enabled Information/Influence Warfare and Manipulation', *SSRN* (23 May 2019), at 7–10.

[42] Chesney and Citron, *supra* note 37, at 28.

[43] Vaidhyanathan, *supra* note 36, at 162.

[44] *Ibid.*, at 163.

hidden from public view and, therefore, less accountable to the media and general public. As a result, political campaigns may feel empowered to make contradictory promises to distinct groups of prospective voters or use microtargeting to deter or suppress particular groups from voting.[45]

Most concerningly, social media platforms have sometimes been relied upon to promote violence against particular individuals and groups. Facebook, for example, has enabled the spread of hate and incitement of violence against Rohingya Muslims in Myanmar,[46] campaigns of harassment and threats against individuals critical of the 'drug war' waged by President Duterte of the Philippines[47] and the exacerbation of divisions between Buddhists and Muslims in Sri Lanka.[48]

Given the growing influence of today's leading social media companies over online content and the rising anxieties that have accompanied their content moderation practices, it is pertinent to ask what state responsibilities arise under HRL with respect to the governance of freedom of expression online. To examine this question, this section draws a distinction between state responsibilities that arise with respect to *liability-driven* content moderation that social media companies are incentivized to undertake in accordance with formal state regulation and state responsibilities that arise with respect to *context-driven* content moderation that social media companies undertake in accordance with the functions and culture of their platforms as well as in response to informal pressures exerted by actors as diverse as states, employees, shareholders, advertisers, mass media organizations, civil society groups and general platform users.[49]

A *Liability-Driven Content Moderation*

Prior to the social media age, one of the greatest threats to freedom of expression was the state's capacity to use criminal law and other coercive measures to *directly* regulate speakers and publishers.[50] With the rise of social media platforms, states have increasingly sought to regulate speakers *indirectly* by relying on a variety of measures to influence the content moderation practices of social media companies.[51] In terms of formal legislation, states predominantly rely on two regulatory mechanisms for this purpose: first, *content restriction laws*, which define categories of content that are illegal in particular domestic and regional contexts and, second, *intermediary liability laws*, which establish the conditions under which intermediaries, including social media companies, may be held liable for unlawful content generated by their users.

45 Borgesius *et al.*, *supra* note 36, at 87.

46 S. Stecklow, 'Hatebook: Why Facebook Is Losing the War on Hate Speech in Myanmar', *Reuters* (15 August 2018).

47 E. Johnson, 'Memo from a "Facebook Nation" to Mark Zuckerberg: You Moved Fast and Broke Our Country', *Recode* (26 November 2018).

48 A. Taub and M. Fisher, 'Where Countries Are Tinderboxes and Facebook Is a Match', *New York Times* (21 April 2018).

49 Jørgensen and Pedersen, *supra* note 11, at 183.

50 Balkin, *supra* note 10, at 1174 (referring to this as 'old school speech regulation').

51 *Ibid.*, at 1176 (referring to this as 'collateral censorship').

Against this background, this section surfaces the contestability of HRL in defining the requirements and circumscribing the limits of these laws and reveals the significance of prioritizing a marketized or structural understanding of HRL for the governance of content on social media platforms.

1 Content Restriction Laws

Turning first to the governance of content restriction laws, HRL performs a dual role, on the one hand mandating that certain forms of expression be prohibited, and on the other hand circumscribing the types of content that may be restricted by states. Under HRL, there are certain types of expression that states are exceptionally obliged to prohibit and refrain from spreading. Article 20(2) of the International Covenant on Civil and Political Rights (ICCPR), for example, provides that states are required to prohibit 'any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence'.[52] In addition, according to the United Nations (UN) special rapporteur on freedom of expression, contemporary HRL offers broader protection against discriminatory hate speech beyond 'national, racial or religious hatred', extending to adverse actions on grounds of 'race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, including indigenous origin or identity, disability, migrant or refugee status, sexual orientation, gender identity or intersex status'.[53] While clear in the abstract, Article 20(2) has given rise to definitional challenges in practice, with judicial guidance circumscribing the types of expression that fall within the prohibition not always clear or consistent.[54] Moreover, the most detailed guidance on Article 20(2) to date – the Rabat Plan of Action, adopted by a high-level group of human rights experts – suggests that states should consider six factors for the purpose of identifying expressions that should be criminally prohibited, an inevitably complex analysis that still leaves room for uncertainty and abuse.[55]

Beyond mandating the restriction of particular categories of content, HRL also limits whether and when different types of content may be restricted by a state. Whenever a state restricts speech, including restrictions pursuant to Article 20(2) of the ICCPR, the UN Human Rights Committee (HRC) has confirmed that the state must adhere to the standards elaborated in Article 19(3) of the ICCPR, which only permits interferences with freedom of expression if restrictions are prescribed by law and necessary for a limited number of legitimate aims – namely, respect for the rights or reputations of others, national security, public order, public health or public morals.[56]

[52] International Covenant on Civil and Political Rights (ICCPR) 1966, 999 UNTS 171, Art. 20(2).

[53] Report of UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN Doc. A/74/486 (9 October 2019), para. 9.

[54] Clooney and Webb, 'The Right to Insult in International Law', 48 *Columbia Human Rights Law Review* (2017) 1, at 38–47.

[55] Rabat Plan of Action, UN Doc. A/HRC/22/17/Add.4 (5 October 2012), Appendix, para. 29.

[56] United Nations Human Rights Committee (UN HRC), General Comment no. 34: Article 19: Freedom of Opinion and Expression, UN Doc. CCPR/C/GC/34 (12 September 2011), para. 52.

Reflecting on how these criteria have been understood in practice, a number of general points emerge.

First, it is difficult to envisage circumstances where generic bans of particular platforms would be deemed necessary under HRL given the significant collateral effects of such measures on the freedom of expression of Internet users.[57] Second, blanket bans of disinformation or untruthful expression generally lack sufficient precision to be compatible with the legality test under Article 19(3) of the ICCPR and also fall foul of the necessity test, bearing in mind that it is well established under HRL that the right to freedom of expression is not limited to 'correct' statements and protects information and ideas that may shock, offend or disturb.[58] Third, since the online speech environment typically varies from state to state, the permissibility under HRL of speech restrictions imposed by a state will also vary to a certain degree according to a contextually informed assessment of the criteria in Article 19(3) of the ICCPR.[59] Fourth, the necessity of a particular restriction will generally depend on a contextually informed assessment of different factors, including the type of speech affected and the measure's proportionality in achieving a legitimate aim.[60] In practice, HRL affords different degrees of protection to different categories of expression. For example, while a high level of protection has typically been afforded to political discourse,[61] states generally have greater leeway in responding to 'gratuitously offensive' speech acts.[62] Beyond examining the type of speech involved, HRL also requires consideration of whether a less far-reaching measure could have been relied upon to achieve the legitimate aim.[63] For example, while states are generally required to prohibit disinformation that amounts to incitement to violence, other forms of false or inflammatory speech may only justify a less restrictive response such as encouraging more speech – whether aimed at promoting diversity and understanding or empowering minorities.[64]

Beyond these general points, however, the open textured nature of the right to freedom of expression has rendered it compatible with different approaches to content restriction laws imposed by states – some more aligned with a market-friendly understanding of HRL and others leaning towards a more structural vision of HRL. This may be illustrated by contrasting the majority judgment and one of the joint

[57] *Ibid.*, para. 43; ECtHR, *Cengiz and Others v. Turkey*, Appl. nos 48226/10 and 14027/11, Judgment of 1 December 2015, paras 47–67. All ECtHR decisions are available at http://hudoc.echr.coe.int/.

[58] Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda, Doc. FOM.GAL/3/17 (3 March 2017), para. 2(a).

[59] M. Milanovic, 'Viral Misinformation and the Freedom of Expression: Part II', *EJIL:Talk!* (13 April 2020).

[60] Leerssen, 'Cut Out by the Middle Man: The Free Speech Implications of Social Network Blocking and Banning in the EU', 6 *Journal of Intellectual Property, Information Technology and E-Commerce Law* (*JIPITEC*) (2015) 99, at 102.

[61] HRC General Comment no. 34, *supra* note 56, para. 38; ECtHR, *Animal Defenders International v. United Kingdom*, Appl. no. 48876/08, Judgment of 22 April 2013, paras 102–104.

[62] ECtHR, *Otto-Preminger-Institut v. Austria*, Appl. no. 13470/87, Judgment of 20 September 1994, para. 49.

[63] HRC General Comment no. 34, *supra* note 56, para. 34.

[64] Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN Doc. A/66/290 (10 August 2011), para. 41.

dissenting opinions in the case of *Animal Defenders International v. United Kingdom*, which concerned the compatibility of a United Kingdom (UK) ban on paid political advertising transmitted via television and radio with the right to freedom of expression under Article 10 of the European Convention on Human Rights (ECHR).[65]

In their judgment, a narrow majority of nine judges upheld the ban by analysing it from a structural HRL perspective. In particular, the majority focused their assessment on the justifications put forward by the UK for imposing 'general measures' – namely, measures that apply to predefined situations regardless of the individual facts of a particular case – pursuant to the state's positive obligation 'to intervene to guarantee effective pluralism in the audiovisual sector' and to ensure 'the maintenance of a free and pluralist debate on matters of public interest, and more generally, contributing to the democratic process'.[66] Amongst a range of factors examined in the judgment, the majority placed particular weight on 'the quality of the parliamentary and judicial review of the necessity of the measure', the fact that the prohibition was 'specifically circumscribed' to paid political advertising on broadcast media whose societal influence remained both 'immediate and powerful', as well as the risk of abuse, uncertainty, litigation, expense, delay, discrimination and arbitrariness if the ban were to be relaxed.[67] Most importantly, the majority took seriously the political economy of political speech and the state's desire 'to protect the democratic debate and process from distortion by powerful financial groups with advantageous access to influential media' who might 'obtain competitive advantage in the area of paid advertising and thereby curtail a free and pluralist debate'.[68]

By contrast, the joint dissenting opinion of Judges Ziemele and colleagues adopted a more marketized perspective of HRL, critiquing the UK's ban as 'an inappropriately assumed positive duty of the State to enable people to impart and receive information', whilst placing greater emphasis on 'the fundamental negative obligation of the State to abstain from interfering'.[69] Focusing more narrowly and individualistically on the particular situation of the applicant in the case, the dissenting judges dismissed concerns that 'powerful groups will invariably hamper the receipt of information by a one-sided information overload' and insisted that a robust democracy would not benefit from the 'benevolent silencing of all voices' or 'well-intentioned paternalism'.[70] Had this opposing understanding of freedom of expression prevailed, the door would arguably have been opened to the higher spending campaigns commonly seen in the USA, whose First Amendment jurisprudence has followed a similarly libertarian approach.[71]

[65] Convention for the Protection of Human Rights and Fundamental Freedoms 1950, 213 UNTS 222.

[66] *Animal Defenders International*, *supra* note 61, paras 106, 111–112.

[67] *Ibid.*, paras 106–125.

[68] *Ibid.*, para. 112.

[69] *Ibid.*, para. 12, Joint Dissenting Opinion of Judges Ziemele *et al.*

[70] *Ibid.*, paras 12, 14.

[71] See similarly Rowbottom, '*Animal Defenders International:* Speech, Spending, and a Change of Direction in Strasbourg', 5 *Journal of Media Law* (2013) 1, 5–6.

Applied to the social media context, the opposing conceptions of HRL in *Animal Defenders International* are potentially very significant, affecting, for example, whether the door is left open for state regulation of online political advertising and microtargeting practices where necessary to protect democratic debate from distortion by powerful groups in society.[72] Although the precise form of regulation will always require assessment in light of the specificities of particular societal contexts, it is suggested that the structural understanding of HRL is to be preferred in light of its openness to positive state intervention where necessary to correct the market-driven distortion of democratic processes that may arise on social media platforms.

2 Intermediary Liability Laws

In the social media governance context, content restriction laws are closely tied to intermediary liability laws, which circumscribe the liability that can be imposed on intermediaries – including social media companies – for unlawful content posted by their users. By defining the conditions that social media companies must satisfy in order to benefit from immunity from legal claims concerning unlawful user-generated content, intermediary liability laws shape platform incentives to respond to potentially unlawful content and to protect the freedom of expression of their users. Given their potential impact on the freedom of expression of platform users, it is pertinent to examine their permissibility under HRL. To date, however, human rights authorities and experts have offered conflicting guidance concerning the types of safeguards that must be established for intermediary liability laws to comply with HRL.

On the one hand, some authorities appear to have prioritized a more marketized conception of HRL, one that focuses narrowly on preventing the harms associated with illegal speech shared on online platforms at the expense of reflecting on how intermediary liability laws may structurally incentivize platforms to 'over-remove' by taking down lawful content in order to avoid liability and/or deter innovators from establishing new platforms in the first place.[73] The European Court of Human Rights (ECtHR), for example, has been relatively permissive of intermediary liability laws, albeit in the narrow context of examining the liability of online news portals for comments posted by readers below their articles. In *Delfi AS v. Estonia*, the Grand Chamber concluded that states could impose liability on large commercially run online news portals for failing to take measures to remove 'clearly unlawful' comments amounting to hate speech and direct threats to the physical integrity of individuals without delay '*even without notice* from the alleged victim or from third parties'.[74] By legitimating the imposition of liability on a company for failing to proactively monitor the content on its website, the Court seems to have neglected to consider the significant collateral effects that such an intermediary liability framework could have on the free

[72] Dobber *et al.*, 'The Regulation of Online Political Micro-targeting in Europe', 8 *Internet Policy Review* (2019) 1, at 11.

[73] Van Hoboken and Keller, 'Design Principles for Intermediary Liability Laws', *Transatlantic Working Group* (8 October 2019), at 4.

[74] ECtHR, *Delfi AS v. Estonia*, Appl. no. 64569/09, Judgment of 16 June 2015, para. 159 (emphasis added).

speech rights of users. As Judges Sajó and Tsotsoria explain in their joint dissenting opinion, such a regime effectively requires online news portals to proactively monitor all comments from the moment they are posted in order to avoid liability, with the consequence that portals 'will have considerable incentives to discontinue offering a comments feature'.[75]

These concerns were acknowledged in the subsequent case of *MTE v. Hungary*, which attempted to confine the proactive monitoring obligation in *Delfi* to 'clearly unlawful' user comments.[76] For other types of content, the Court concluded that a notice and takedown model – where content is reviewed and potentially removed by social media platforms following notice – accompanied by effective procedures allowing for rapid response would suffice.[77] Yet, in attempting to restrict the effects of the *Delfi AS* ruling, the Court in *MTE* failed to consider the all-or-nothing nature of online monitoring.[78] In practice, for an online news portal to avoid liability for user-generated hate speech under the intermediary liability framework accepted in *Delfi AS*, it will need to monitor *all* comments.[79] As Daphne Keller explains, '[u]nder a *Delfi*/*MTE* rule, tech platforms would still go looking for hate speech, find other potentially unlawful content, and presumably remove it – with precisely the "foreseeable negative consequences on the comment environment of an Internet portal" and "chilling effect on the freedom of expression on the Internet" that the Court identified and tried to avoid'.[80] With this in mind, arguably the most important finding across these two cases is the Court's conclusion in *Delfi AS* that the case was concerned solely with large professionally run online news portals and not with other Internet fora where user-generated comments can be disseminated.[81] It remains to be seen whether the ECtHR will adopt a comparable approach and reach similar conclusions with respect to the liability of social media companies.

By contrast, other authorities have adopted a more structural understanding of HRL in their examination of intermediary liability laws, taking a more holistic view of the systemic effects that such laws may have on the online speech environment. According to the Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, for example, a model of strict liability according to which intermediaries are held liable for unlawful content generated by third parties without notice is incompatible with the right to freedom of expression because it creates 'strong incentives for the private censorship of a wide range of legitimate expression'.[82] The Office of the Special Rapporteur also concluded that notice

[75] *Ibid.*, para. 1, Joint Dissenting Opinion of Judges Sajó and Tsotsoria.

[76] ECtHR, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu ZRT v. Hungary*, Appl. no. 22947/13, Judgment of 2 February 2016.

[77] *Ibid.*, para. 91.

[78] Keller, 'Policing Online Comments in Europe: New Human Rights Case Law in the Real World', *Center for Internet and Society* (12 April 2016).

[79] P.-J. Ombelet and A. Kuczerawy, 'Delfi Revisited: The MTE-Index.hu v. Hungary Case', *LSE Media Policy Project Blog* (19 February 2016).

[80] Keller, *supra* note 78.

[81] *Delfi AS v. Estonia*, *supra* note 74, para. 116.

[82] Office of the Special Rapporteur for Freedom of Expression, Inter-American Commission on Human Rights, 'Freedom of Expression and the Internet', Doc. OEA/Ser.L/V/II.CIDH/RELE/INF/11/13 (31 December 2013), paras 95–103.

and takedown regimes would only be compatible with the right to freedom of expression to the extent that they incorporate sufficient judicial safeguards.[83] A similar approach was adopted by Argentina's Supreme Court in the landmark *Belén Rodriguez* case.[84] Reasoning from constitutional and human rights sources, the Supreme Court held that search engines could be found liable for third-party content only if they had actual knowledge of illicit content and failed to take remedial steps. In specifying the meaning of 'actual knowledge', the Court distinguished between ostensible infringing content, such as child exploitation material, for which private notification would suffice, and other content for which judicial notification would be necessary.[85]

These latter authorities recognize that intermediary liability laws unaccompanied by adequate safeguards are incompatible with HRL because of their potential to generate overbroad and disproportionate effects on the freedom of expression of platform users. In this regard, while consensus is yet to emerge on the precise constellation of safeguards that may be deemed adequate according to more structurally oriented perspectives of HRL, a number of general principles are beginning to emerge.[86]

For example, it is increasingly recognized that the imposition of proactive monitoring and filtering obligations on platforms is disproportionate for all but the most manifestly unlawful and easy to adjudicate categories of content such as child exploitation material, particularly in light of the limited ability of technical filters to assess context or adapt to the coded and evolving meaning of language.[87] In addition, in order to guard against incentivizing the over-removal of lawful content, there is growing acknowledgement that intermediary liability laws must incorporate a package of public and private due process and accountability safeguards, making platform immunity for most categories of unlawful content contingent on the receipt of a notice issued by a judicial or quasi-judicial public authority as well as the provision of private procedural protections for speakers when action is taken against their content.[88] Importantly,

[83] *Ibid.*, paras 104–108. See similarly Joint Declaration on Fake News, *supra* note 58, para. 1(d); Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation (24 March 2015).

[84] Supreme Court (Argentina), *Rodriguez M. Belén c/Google Inc. s/daños y perjudicios*, Judgment R.522.XLIX (28 October 2014).

[85] *Ibid.*, para. 18; see also Supreme Court (India), *Shreya Singhal v. Union of India*, Judgment no. 1672/2012 (24 March 2015), para. 117 (defining 'actual knowledge' as notice from 'a court order').

[86] See generally L. Belli and N. Zingales (eds), *Platform Regulations: How Platforms Are Regulated and How They Regulate Us* (2017); Council of Europe, Recommendation CM/Rec (2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries (7 March 2018); Land, 'Against Privatized Censorship: Proposals for Responsible Delegation', 60 *Virginia Journal of International Law* (2020) 363; Transatlantic Working Group (TWG), *Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online* (2020); Global Network Initiative, *Content Regulation and Human Rights* (2020); Access Now, *Access Now's Position on the Digital Services Act Package* (2020).

[87] See, e.g., Council of Europe, *supra* note 86, para. 1.3.8; Land, *supra* note 86, at 418–425; Global Network Initiative, *supra* note 86, at 22; see also Oliva, 'Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression', 20 *Human Rights Law Review* (2020) 607.

[88] See, e.g., Land, *supra* note 86, at 429; Global Network Initiative, *supra* note 86, at 13–14, 24. One option for ensuring adequate safeguards in content moderation processes would be to rely on 'notice-and-notice' procedures for at least some categories of less serious illegal content (for example, civil claims relating to copyright, defamation and privacy), on which see generally Article 19, *Internet Intermediaries: Dilemma of Liability* (2013).

taking account of the scale of content disseminated across online platforms, there is also growing recognition that systemic conceptions of due process and accountability are required in the social media context – for example, by differentiating levels of due process depending on the type of content under review and making use of targeted, rather than comprehensive, forms of accountability such as audits focused on samples of cases.[89] Beyond due process and accountability, it is also generally accepted that the application of HRL's legality test in the intermediary liability context requires that the categories of unlawful content for which social media companies may potentially be held liable must be defined with sufficient precision and clarity in order to minimize the collateral removal of lawful speech and guard against 'censorship creep' whereby ambiguity leads to ever-expanding categories of content being restricted in practice.[90]

In terms of the extraterritorial scope of intermediary liability laws, there is increasing support for the proposition that states should only exceptionally require social media platforms to comply with domestic content restriction requirements on a global basis. Whether unilaterally imposed global speech restrictions are normatively desirable depends on what is being restricted and whether the order in question includes onerous requirements to remove additional content – for example, 'identical' or 'equivalent' posts – beyond the post initially identified.[91] Significantly, rules of jurisdiction, conflict of laws and comity will generally not prevent platforms from voluntarily enforcing cross-border speech restrictions.[92] As Jennifer Daskal has observed, '[a]bsent some sort of must-carry obligation, takedown and delisting obligations merely compel companies to do something that they can do voluntarily'.[93] As such, in the absence of resistance from the platforms themselves, the risk arises of a race to the bottom whereby the most censorship-prone states are able to define the boundaries of freedom of expression online.[94]

However, in *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, European Union (EU) Advocate General Szpunar explained that, in order to respect widely recognized fundamental rights, Member State courts should 'as far as possible' limit the extraterritorial effects of their injunctions concerning harm to private life and personality rights, refraining from going beyond 'what is necessary to achieve the protection of the injured person' and 'in an appropriate case' order that access to that information be disabled through geo-blocking.[95] The Court of Justice of the European Union

[89] See, e.g., douek, 'Verified Accountability: Self-Regulation of Content Moderation as an Answer to the Special Problems of Speech Regulation', Aegis Series Paper no. 1903 (2019), at 8–11; Land, *supra* note 86, at 429.

[90] Citron, 'Extremist Speech, Compelled Conformity, and Censorship Creep', 93 *Notre Dame Law Review* (2018) 1035, at 1051.

[91] See generally Daskal, 'Speech across Borders', 105 *Virginia Law Review* (2019) 1605, at 1650ff.

[92] Keller, 'Who Do You Sue? State and Platform Hybrid Power over Online Speech', Aegis Series Paper no. 1902 (2019), at 9.

[93] Daskal, *supra* note 91, at 1652.

[94] Balkin, *supra* note 10, at 1206. See, however, Woods, 'Litigating Data Sovereignty', 128 *Yale Law Journal* (2018) 328, at 391–393 (explaining how comity and public-policy exceptions to general comity principles may limit the reach or strength of global injunctions that violate fundamental values).

[95] Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* (EU:C:2019:458), para. 100, Opinion of Advocate General Szpunar.

(CJEU) has also emphasized that Member State courts should ensure extraterritorial injunctions are 'consistent with the rules applicable at the international level',[96] only ordering global injunctions 'where appropriate' after weighing the human rights at issue against each other.[97] While the precise circumstances when a global injunction may be deemed necessary and proportionate in accordance with HRL remain to be clarified through case law, Advocate General Szpunar's opinion implies that the extraterritorial application of speech restrictions should be viewed as an exceptional measure that only applies where there is a sufficiently strong interest at stake.[98]

Finally, in terms of the nature of the obligations placed on intermediaries, it is increasingly accepted that intermediary liability laws should be 'graduated and differentiated' based on the size of a platform so as to avoid over-burdening smaller start-ups and further entrenching the market dominance of today's largest social media companies.[99] Moreover, given the inevitability of errors in operationalizing content moderation at scale, there is also growing recognition that platform immunity should be contingent on systemic deficiencies rather than individual errors within platform moderation systems.[100] In this regard, it is notable that, while Germany's *NetzDG* legislation has been justifiably criticized for requiring the removal of vaguely defined categories of unlawful content within unreasonably narrow time frames under threat of substantial financial penalties,[101] it nonetheless illustrates an intermediary liability regime that is differentiated (confined to social media companies with at least 2 million users within Germany) and systemic (applying penalties only to systematic and persistent failures in the complaints management systems that platforms are required to establish).[102]

Each of these regulatory tools are illustrations of what Joris van Hoboken and Daphne Keller refer to as the 'dials and knobs' available to lawmakers that can alter the incentive structure of intermediary liability laws.[103] Of course, in practice, the devil is in the detail, and a persistent challenge in this regulatory context remains the dearth of platform transparency concerning the operation of their content moderation systems combined with the weak empirical basis that currently exists to inform the design of intermediary liability laws.[104] With this in mind, a prerequisite for operationalizing a more structural HRL approach to the design of intermediary liability

[96] Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited* (EU:C:2019:821), paras 51–52.

[97] Case C-507/17, *Google v. Commission Nationale de l'Informatique et des Libertés (CNIL)* (EU:C:2019:772), para. 72.

[98] See similarly, Daskal, *supra* note 85, at 1655–1657 (proposing a rebuttable presumption in favour of geographical segmentation).

[99] See, e.g., Council of Europe, *supra* note 86, para. 1.3.9; Global Network Initiative, *supra* note 86, at 20.

[100] See, e.g., Global Network Initiative, *supra* note 86, at 21; douek, 'Governing Online Speech', *supra* note 22, at 46–51, 64–67.

[101] See generally Schulz, 'Roles and Responsibilities of Information Intermediaries: Fighting Misinformation as a Test Case for a Human Rights-Respecting Governance of Social Media Platforms', Aegis Series Paper no. 1904 (2019), at 10–14.

[102] See similarly Theil, 'The New German Social Media Law: A Risk Worth Taking?', *UK Human Rights Blog* (19 February 2018).

[103] Van Hoboken and Keller, *supra* note 73, at 4; see also Roberts, 'Digital Detritus: "Error" and the Logic of Opacity in Social Media Content Moderation', *First Monday* (5 March 2018).

[104] See generally Keller and Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation', in N. Persily and J.A. Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (2020) 220.

laws is for states to mandate greater and more granular transparency from social media platforms, whilst also monitoring how particular design choices implemented through their intermediary liability laws affect the online information ecosystem in particular societal contexts. Only with enhanced transparency can states begin to ensure that the design of their intermediary liability laws is based on, and evolves in response to, concrete evidence concerning the effects of platform policies and intermediary liability laws on the online environment as a whole.[105]

B *Context-Driven Content Moderation*

In practice, content moderation on social media platforms is incentivized not only by formal state legislation but also by a broader set of contextual factors. States, for example, may exert influence over the flow of information online and the moderation practices of social media companies by conducting their own information campaigns on their platforms. In addition, any state with sufficient influence – for example, because it controls access to a commercially valuable market – may exert pressure over social media companies by taking advantage of more informal regulatory techniques.[106] Examples include deploying Internet referral units that flag user-generated content to platforms for review against their terms of service; initiating various forms of informal cooperation through which social media companies undertake to 'voluntarily' respond to removal requests concerning particular categories of content within specific time-frames or face the prospect of future formal regulation; and jawboning through public appeals by government officials urging social media companies to improve their capabilities for addressing particular categories of content.[107]

Significantly, by influencing the formulation of platforms' terms of service and community standards that typically apply uniformly around the world, these more informal pressures have the potential to affect the restriction of speech on a global scale. In addition, where effective, informal pressures enable states to influence platform moderation practices whilst circumventing the scrutiny and accountability that typically accompanies more formal channels such as domestic judicial processes.[108] By way of illustration, social media companies appear to have implemented opaque and unaccountable forms of cross-platform collaboration to remove content or actors from their sites as a means of averting the prospect of future state regulation.[109] Notably, although these so-called 'content cartels' initially arose in response to rising governmental concerns over the spread of terrorist content, they now appear to be expanding to encompass ever-wider categories of content.[110]

[105] See similarly douek, 'Governing Online Speech', *supra* note 22, at 58; TWG, *supra* note 86, at 22–25; see also Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems', 11 *EJLT* (2020).

[106] Keller, *supra* note 92, at 7.

[107] Sander, *supra* note 24, at 951–952.

[108] Article 19, *Side-stepping Rights: Regulating Speech by Contract* (2018), at 16–17.

[109] Radsch, 'GIFCT: Possible the Most Important Acronym You've Never Heard Of', *Just Security* (30 September 2020).

[110] douek, 'The Rise of Content Cartels', *Knight First Amendment Institute* (2020).

Finally, beyond pressures exerted by states, platform moderation is also influenced to varying degrees by commercial and reputational concerns, including the particular functions and culture of a platform, as well as the interests of advertisers, employees, shareholders, mass media organizations, civil society groups and general users.[111] In practice, these context-driven moderation practices raise a number of thorny questions under HRL, the resolution of which will often depend on whether a marketized or structural conception of HRL is prioritized in practice.

Adopting a marketized conception of HRL that places emphasis on the negative obligation of states to refrain from unjustifiably interfering with the right to freedom of expression, it is only state-sponsored information campaigns that fall squarely within HRL's regulatory purview. It is clear, for example, that such campaigns must not promote content that states are exceptionally required to prohibit under HRL, such as the various forms of discriminatory hate speech prohibited by Article 20(2) of the ICCPR. In addition, states that systematically and surreptitiously flood online platforms with false or inflammatory content may violate the right of individuals to seek and receive information under HRL, particularly where the result is to crowd out accurate information and silence legitimate debate.[112] Finally, reference may also be made to the Committee on Economic, Social and Cultural Rights, which has confirmed that 'the deliberate … misrepresentation of information vital to health protection or treatment' where 'likely to result in bodily harm, unnecessary morbidity and preventable mortality' constitutes a violation of the state's obligation to respect the right to health under Article 12 of the International Covenant on Economic, Social and Cultural Rights.[113]

Yet, although these provisions are clear in the abstract, their application in the social media context is confronted by at least four challenges: first, there is the *definitional* challenge of determining whether an information operation relies upon categories of expression prohibited under HRL, an assessment that will generally be highly context dependent and difficult to conduct given the speed, scale and linguistic diversity of communication on social media platforms;[114] second, there is the *threshold* challenge of determining the precise threshold that must be crossed for the practice of flooding a platform with false or inflammatory content to constitute a violation of the right to seek and receive information under HRL; third, there is the *enforcement* challenge of attributing information campaigns to states, which may attempt to hide their online identity; and, finally, there is the *extraterritoriality* challenge of determining the extent to which HRL applies to cross-border state-sponsored information operations (although, as detailed below in the context of discussing state-sponsored cyber

[111] Leerssen, *supra* note 60, at 111–113.

[112] Milanovic, *supra* note 59.

[113] International Covenant on Economic, Social and Cultural Rights, 1966, 993 UNTS 3, para. 50, cited in Milanovic, *supra* note 59.

[114] See generally Land and Wilson, 'Hate Speech on Social Media: Towards a Context-Specific Content Moderation Policy', 52 *Connecticut Law Review* (2020) 1029.

surveillance operations, there is increasing support from different human rights authorities that suggests this challenge is surmountable).

Beyond these challenges, adhering to a narrow marketized conception of HRL also risks other contextual pressures falling beyond the regulatory scope of HRL. For example, it is at the very least uncertain at what stage restrictions on the freedom of expression of platform users may be attributed to a state when implemented via 'voluntary' measures adopted by social media companies in response to informal governmental pressures.[115] Similarly, a marketized conception of HRL would generally afford states a wide margin of appreciation for determining how to strike a fair balance between the freedom of a social media company to conduct its business (including in response to commercial and reputational concerns) and the right to freedom of expression of platform users.[116]

By contrast, a more structural understanding of HRL would place greater emphasis on developing positive obligations of states that respond to the systemic accountability deficits that pervade the social media ecosystem. In the content moderation context, positive obligations under HRL may arise from at least two bases. First, to the extent that today's market-dominant social media companies may be considered to be exercising an inherent governmental function through their regulation of online speech,[117] a positive obligation requiring states to ensure that such privatization does not undermine HRL arises. As the Inter-American Court of Human Rights confirmed in *Ximenes-Lopes v. Brazil*, delegating public services to private institutions 'requires as an essential element the responsibility of the States to supervise their performance in order to guarantee the effective protection of the human rights of the individual under the jurisdiction thereof and the rendering of such services to the population on the basis of non-discrimination and as effectively as possible'.[118]

Second, states are also under a positive obligation to ensure that persons within their territory and/or jurisdiction are protected from acts of private actors that would impair their enjoyment of the right to freedom of expression.[119] According to the ECtHR in *Dink v. Turkey*, for example, states are under a positive obligation 'to create a favourable environment for participation in public debate by everyone and to enable the expression of opinions and ideas without fear'.[120] This statement is complemented by the UN HRC's confirmation that the right to political participation under Article 25 of the ICCPR requires that citizens 'must be free to vote *without undue influence or coercion of any kind* which may distort or inhibit the free expression of the elector's will' and 'should be able to form opinions independently, *free of violence or threat of violence,*

[115] Jørgensen and Pedersen, *supra* note 11, at 186–187. See, however, Land, *supra* note 80, at 396–409.

[116] Jørgensen and Pedersen, *supra* note 11, at 183–184.

[117] See generally Belli, Francisco and Zingales, 'Law of the Land or Law of the Platform? Beware the Privatisation of Regulation and Police', in Belli and Zingales, *supra* note 86; Land, *supra* note 86, at 406, 413.

[118] IACtHR, *Ximenes-Lopes v Brazil*, Judgment (Merits, Reparations and Costs), 4 July 2006, para. 96.

[119] HRC General Comment no. 34, *supra* note 56, para. 7.

[120] ECtHR, *Dink v. Turkey*, Appl. nos 2668/07, 6102/08, 30079/08, 7072/09 and 7124/09, Judgment of 14 September 2010, para. 137 (emphasis added; author's translation).

compulsion, inducement or manipulative interference of any kind'.[121] Importantly, this positive obligation is not confined to electoral periods. In *Animal Defenders International*, for example, the ECtHR observed that 'while the risk to pluralist public debates, elections and the democratic process would evidently be more acute during an electoral period, … the democratic process is a continuing one to be nurtured at all times by a free and pluralist public debate'.[122]

Viewing the contextual pressures that influence content moderation through the prism of a state's positive obligations under HRL, the pertinent task becomes identifying what measures and safeguards states must establish to ensure a favourable online environment for participation in public debate by everyone. In general, human rights courts have refrained from requiring specific types of intervention to satisfy positive obligations concerning the right to freedom of expression. In *Verein gegen Tierfabriken v. Switzerland*, for example, the ECtHR concluded that 'it is not the Court's task to indicate which means a State should utilise in order to perform its obligations under the Convention', its role confined to determining 'whether the Contracting States have achieved the result called for by the Convention'.[123] Similarly, in *Animal Defenders International*, the Court emphasized that 'there is a wealth of historical, cultural and political differences within Europe so that it is for each State to mould its own democratic vision', and observed that 'the legislative and judicial authorities are best placed to assess the particular difficulties in safeguarding the democratic order in their State'.[124] Nonetheless, it is possible to identify a number of guidelines for state intervention in this context rooted in a more structural conception of HRL.

In terms of pressures to remove content from social media platforms, any regulatory framework should clearly distinguish between unlawful speech and 'lawful but harmful' content.[125] In terms of unlawful speech, state pressure to incentivize the removal of such content should be confined to formal intermediary liability frameworks that incorporate adequate safeguards against arbitrary and discriminatory removal as already outlined. This means that states should refrain from establishing and utilizing extrajudicial mechanisms to restrict content such as Internet referral units or 'voluntary' cooperation agreements that lack the necessary transparency, definitional specificity, due process, accountability or oversight to protect against arbitrary and discriminatory content removal practices.[126] Equally, states should ensure that any cross-platform initiatives that seek to centralize content moderation through collaborative vehicles are subject to robust forms of oversight – such as independent audits – and incorporate adequate protections against abuse.[127]

[121] UN HRC, General Comment no. 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service (Art. 25), UN Doc. CCPR/C/21/Rev.1/Add.7 (12 July 1996), para. 19 (emphasis added).

[122] *Animal Defenders International*, *supra* note 61, para. 111.

[123] ECtHR, *Verein gegen Tierfabriken (VgT) v. Switzerland*, Appl. no. 24699/94, Judgment of 28 June 2001, para. 63.

[124] *Animal Defenders International*, *supra* note 61, para. 111.

[125] TWG, *supra* note 86, at 16.

[126] Manila Principles, *supra* note 83, Principle VI(b).

[127] Radsch, *supra* note 109; douek, *supra* note 110.

In terms of 'lawful but harmful' content, the principles of necessity and proportionality under HRL suggest that state intervention should be minimalist, confined to verifying the systems and processes established by social media platforms to respond to such content.[128] To date, regulatory initiatives in this sphere have been found wanting. With respect to online disinformation, the EU Code of Practice on Disinformation, which constitutes a co-regulatory scheme adopted by online platforms and the advertising industry under the shadow of principles enunciated in the European Commission's communication on tackling online disinformation, elaborates a range of potentially useful policy commitments such as disrupting advertising revenues of accounts that spread disinformation and empowering the research community to monitor online disinformation, but fails to establish measurable objectives, meaningful safeguards against arbitrary interference with freedom of expression, human rights impact assessments or mechanisms to incentivize and verify implementation.[129]

More promising are proposals that adopt a structural human rights lens by seeking to address more systemically how online content is distributed by platform amplification and targeting systems.[130] Access Now, for example, has put forward an approach that seeks to address lawful but harmful content by protecting user choice, enhancing user autonomy and ensuring meaningful public accountability with respect to the open recommender systems utilized by social media platforms.[131] According to this approach, user choice could be protected by ensuring that platform default settings require an 'opt in' to platform personalization systems and by applying proportional sanctions for systemic violations of existing legal frameworks governing data protection, equal treatment and non-discrimination that seek to mitigate engagement-driven human rights abuses. User autonomy could be enhanced by mandating meaningful forms of transparency concerning the algorithmic decision-making that underpins content recommender systems and by requiring platforms to provide greater user control over the operation of such systems – for example, by enabling users to exclude certain content or sources of content from their recommendations. Finally, meaningful public accountability could be facilitated by establishing a robust data access framework that aims to 'allow for research-based policy making and reinforce public scrutiny over gatekeepers' operations that directly impact users' fundamental rights'.[132]

Taken together, the obligations that comprise this approach would need to be applied progressively and pragmatically depending on the size of the social media platform in order to avoid creating entry barriers for new platforms and be supervised by an independent administrative authority, acting in partnership with other regulatory authorities and open to civil society.[133] Moreover, as a growing number of

[128] douek, *supra* note 89 (proposing a model of 'verified accountability').

[129] Article 19, 'EU: New Code of Practice on Disinformation Fails to Provide Clear Commitments, or Protect Fundamental Rights' (9 October 2018); Kuczerawy, 'Fighting Online Disinformation: Did the EU Code of Practice Forget about Freedom of Expression', in E. Kuzelewska *et al.* (eds), *Disinformation and Digital Media as a Challenge for Democracy* (2020) 291.

[130] See, e.g., Maréchal, MacKinnon and Dheere, *supra* note 39.

[131] Access Now, *supra* note 86, ch. 2.

[132] *Ibid.*, at 10.

[133] See similarly French Secretary of State for Digital Affairs, 'Creating a French Framework to Make Social Media Platforms More Accountable: Acting in France with a European Vision' (May 2019), at 17, 22–23.

commentators have emphasized, it is vital that these types of proposals 'intentionally center the experiences, expertise and voices of marginalized communities and critical activists, particularly from the Global South …, as well as informed academic research that empirically assesses the impacts of platform moderation'.[134] Bearing this in mind, Access Now's proposal has the advantage of being: *transparent* and *inclusive*, recognizing the importance of evidence-based policy-making informed by public scrutiny from a plurality of stakeholders including civil society and independent researchers;[135] *flexible* and *minimalist*, focused on regulating platform systems and processes rather than substantive rules for online speech;[136] and *measured* and *verifiable*, through the establishment of a supervisory system that is proportionate and mitigates against risks to freedom of expression – for example, by only subjecting platforms to proportional fines if they first fail to adhere to a prohibition on the deployment of their recommender systems when systemic violations of data protection, equal treatment and non-discrimination frameworks have been identified.[137]

Significantly, there are signs that at least some aspects of this approach are gaining traction amongst policy-makers. The recently published proposal for a Digital Services Act (DSA) by the European Commission, for example, includes suggestions to impose a number of obligations on 'very large online platforms' including transparency requirements for their recommender and advertising systems, user controls over the main parameters of recommender systems including at least one option that is not based on profiling, a data access framework and independent audits to monitor compliance.[138] While the current draft of the DSA requires further work if it is to become more fully aligned with a structural human rights approach,[139] these provisions could provide a basis to move in such a direction.

At the same time, while the preceding suggestions offer promising avenues for making the systems and processes of social media platforms more accountable, without more, they neglect to address the freedom of expression concerns that stem from the market dominance of a handful of social media companies in various societies around the world.[140] One question raised by such dominance is whether states are

[134] Gregory, 'Truth, Lies, and Social Media Accountability in 2021: A WITNESS Perspective on Key Priorities' *WITNESS* (2021).

[135] On the importance of enabling a more pluralist conversation on social media governance, see generally Hamilton, 'Governing the Global Public Square', *Harvard International Law Journal* (forthcoming 2021).

[136] douek, 'Governing Online Speech', *supra* note 22, at 59.

[137] Access Now, *supra* note 86, at 7.

[138] European Commission, Proposal for a Digital Services Act, Doc. COM (2020) 825 final (15 December 2020), ch. 4, s. 4.

[139] For example, the Digital Services Act's proposal that very large online platforms be required to assess and mitigate 'systemic risks' needs further thought given the vagueness of the threshold of 'systemic' risk and uncertainty over what mitigation measures would be considered 'reasonable, proportionate and effective' in practice. See similarly Article 19, 'At a Glance: Does the EU Digital Services Act Protect Freedom of Expression?' (11 February 2021).

[140] See similarly Helberger, 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power', 8 *Digital Journalism* (2020) 842, at 849; Stasi, 'A Capital Riot and Big Tech Takes a Stand: But Is It the One We Want?', *Just Security* (15 January 2021).

under a positive obligation under HRL to compel dominant companies to allow users and at least some categories of lawful content to remain on their platforms (so-called 'must-carry' obligations).[141] Pursuant to the existing jurisprudence of the ECtHR, this question seems to hinge on whether the users in question have access to viable alternative platforms to exercise their right to freedom of expression. The leading authority for this proposition is *Appleby & Others v. United Kingdom*, in which the ECtHR concluded that the right to freedom of expression 'does not bestow any freedom of forum for the exercise of that right' nor does it require 'the automatic creation of rights of entry to private property'.[142] The Court added, however, that it would not exclude that a positive obligation could arise where 'the bar on access to property has the effect of preventing any effective exercise of freedom of expression or it can be said that the essence of the right has been destroyed'.[143]

Applied to the social media context, opinion remains divided as to the implications of the viable alternative platform test. According to Rikke Jørgensen and Anja Pedersen, positive obligations to protect speakers will arise only exceptionally – for example, where a platform 'deprives an online speaker from reaching an audience completely – or deprives an end-user completely from accessing certain content'.[144] Since content banned from Facebook may still be permissible on other platforms, the circumstances that Jørgensen and Pedersen refer to will rarely arise in practice. Paddy Leerssen, by contrast, points to the fact that networks of friends or followers have to be built up over time on platforms, as well as the varying purposes of different platforms, to suggest that social media platforms 'are not necessarily interchangeable, and that end users may lack viable alternatives if removed from a particular service'.[145] This argument is strengthened by the growing concentration of the social media market, with platforms like Facebook and YouTube becoming increasingly dominant channels of communication across the world.[146] Significantly, an indication that the ECtHR may be sympathetic to the latter line of thinking is identifiable in the case of *Cengiz and Others v. Turkey*, where the Court characterized YouTube as 'a unique platform on account of its characteristics, its accessibility and above all its potential impact',[147] adding that the video-sharing platform contained 'specific information of interest to the applicants that is not easily accessible by other means'.[148] These remarks could potentially lay the foundations for the Court

[141] See generally Leerssen, *supra* note 60, at 102–105; Angelopoulos *et al.*, Study on Fundamental Rights Limitations for Online Enforcement through Self-Regulation (2015), at 33–39; Kuczerawy, 'The Power of Positive Thinking: Intermediary Liability and the Effective Enjoyment of the Right to Freedom of Expression', 8 *JIPITEC* (2017) 226, at 229–231; see also Keller, *supra* note 92, at 11–22; Ketteman and Tiedeke, 'Back Up: Can Users Sue Platforms to Reinstate Deleted Content?', 9 *Internet Policy Review* (2020).

[142] ECtHR, *Appleby and Others v. United Kingdom*, Appl. no. 44306/98, Judgment of 6 May 2003, para. 47.

[143] *Ibid.*; see also *VgT v. Switzerland*, *supra* note 123, para. 77; *Animal Defenders International*, *supra* note 61, paras 117, 124.

[144] Jørgensen and Pedersen, *supra* note 11, at 184.

[145] Leerssen, *supra* note 60, at 104.

[146] W. Benedek and M.C. Kettemann, *Freedom of Expression and the Internet* (2013), at 106.

[147] *Cengiz and Others*, *supra* note 57, para. 52.

[148] *Ibid.*, para. 51.

to recognize some form of must-carry obligation, at least for market-dominant social media platforms.[149]

However, even if the ECtHR were to move in this direction, a number of difficult issues would remain to be resolved, including the thorny questions of which categories of speech would fall within the scope of the positive obligation and which platforms – and, more specifically, which of their products and features – would be affected in practice.[150] One emerging source of guidance may be found in recent case law before domestic courts in Germany. As Matthias Ketteman and Anna Tiedeke explain, 'depending on the importance of a communication made (user-side) and the "significant market power" (intermediary side), social network services in Germany face restrictions in limiting access to the platform by suspending users or cancelling profile access contracts via the concept of indirect third-party effect of fundamental rights'.[151] Restrictions may concern the design of terms of service, the interpretation of the terms of service in light of the Basic Law or obligations that platforms are required to take into account such as the equality principle.[152]

Yet, even if a workable solution for recognizing must-carry obligations under HRL were identified, such obligations would still serve to ratify or at the very least neglect to address the dominance of today's leading social media companies rather than offering an avenue for structurally enabling a more diverse and pluralized social media environment. In this regard, it is important to recognize that the freedom of expression concerns raised by the market dominance of particular social media companies extend beyond must-carry issues to what Natali Helberger has termed their 'systemic opinion power' – namely, their capacity to 'create dependencies and influence other players in a democracy' and 'directly and permanently impact the pluralistic public sphere'.[153] In order to tackle these concerns, a structural human rights approach would place greater emphasis on the positive obligation of states to ensure a diverse and pluralistic environment necessary for individuals to effectively exercise their freedom of expression.[154] Such an approach would require states to identify avenues for dispersing the systemic opinion power currently concentrated in today's leading online platforms.[155] To this end, civil society group Article 19, for example, has proposed an unbundling obligation, which would require social media companies to functionally separate their hosting and content moderation services in order to enable competitors to provide competing customized interfaces with bespoke moderation practices on their platforms.[156] While not without technical or

[149] Leerssen, *supra* note 60, at 104–105.

[150] See generally Keller, *supra* note 92, at 13–15.

[151] Ketteman and Tiedeke, *supra* note 141, at 11.

[152] *Ibid.*

[153] Helberger, *supra* note 140, at 846.

[154] See, e.g., *Animal Defenders International*, *supra* note 61, para. 101 (discussing the principles concerning pluralism in the audiovisual media).

[155] See generally Kadri, 'Digital Gatekeepers', *SSRN* (July 2020), at 33–38.

[156] Article 19, Why Decentralisation of Content Moderation Might Be the Best Way to Protect Freedom of Expression Online (30 March 2020); see also Masnick, 'Protocols Not Platforms: A Technological Approach to Free Speech', Knight First Amendment Institute (21 August 2019); Doctorow, 'Adversarial Interoperability', *EFF Deeplinks* (2 October 2019) (emphasis in original); Keller *supra* note 92, at 26–27; J. Cobbe and E. Bietti, 'Rethinking Digital Platforms for the Post-COVID-19 Era', *CIGI* (12 May 2020);

data protection challenges that would need to be met,[157] an unbundling obligation that required platforms to open a moderation application programming interface to potential competitors would potentially undercut the high switching costs that characterize today's market-dominant platforms and help nurture a more pluralized social media landscape.

In practice, of course, nurturing a more pluralized social media landscape is to some extent in tension with subjecting today's largest online platforms to ever-more burdensome requirements in an effort to ensure that their moderation systems are infused with public, rather than purely profit-driven, interests. Going forward, therefore, a central regulatory challenge for those pursuing a structural human rights approach will be to identify the optimal trade-off between these regulatory approaches.

3 Data Surveillance

Closely entwined with the content moderation architecture of social media platforms is an extensive system of data surveillance. To make a profit, social media companies enable individuals and groups to connect and communicate on a global scale in exchange for surveilling their data. Data surveillance is financially lucrative for social media companies, enabling them to algorithmically personalize their platforms in ways that maximize user engagement and monetize user attention through the sale of targeted advertising.[158] The extractive logic of this business model incentivizes social media companies to amass as much data as they can – whether derived from their platforms, elsewhere on the Internet or third parties.[159] Importantly, this business model not only demands that individuals trade their privacy for the ability to communicate online but also generates significant opportunities for abuse of power – whether in the form of manipulative microtargeting, false or inflammatory information operations, addictive platform design or vulnerabilities to third-party access to personal data.[160] In order to examine the responsibilities of states under HRL that arise with respect to the data surveillance ecosystem that underpins the platform economy, this section examines social media companies from two perspectives:[161] first, as *surveillance intermediaries* that stand between the state and user data; and second, as *surveillance principals* that process user data for their own commercial interests.

I. Brown, *Interoperability as a Tool for Competition Regulation* (2020); and I. Brown, *The Technical Components of Interoperability as a Tool for Competition Regulation* (2020). For a sceptical view, see Kwet, 'Fixing Social Media: Toward a Democratic Digital Commons', 5 *Markets, Globalization and Development Review* (2020) 1.

157 Keller, *supra* note 92, at 27; Balkin, 'The Fiduciary Model of Privacy', 134 *Harvard Law Review Forum* (2020) 11, at 22.

158 Balkin, 'Fixing Social Media's Grand Bargain', Aegis Series Paper no. 1814 (2018), at 2–3.

159 Amnesty International, *Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights* (2019), at 8–16.

160 Balkin, *supra* note 158, at 3–5.

161 Rozenshtein, *supra* note 12; Cohen, *supra* note 12, at 191–199.

A *Surveillance Intermediaries*

The sheer scale of data processed on today's largest social media platforms amounts to a potential treasure trove for law enforcement and intelligence agencies. The significance of social media companies as intermediaries for state surveillance first garnered worldwide attention in 2013 when Edward Snowden disclosed details of collaboration between social media companies and the United States National Security Agency as part of its PRISM programme.[162] Since the Snowden disclosures, social media companies have become more resistant to state surveillance efforts, incentivized by a combination of commercial, ideological and security interests.[163] At the same time, however, the surveillance practices of states have expanded, a trend reflected not only in the rising number of government requests for social media user data,[164] but also in the growing number of advanced social media monitoring programs and the booming market for social media surveillance tools.[165]

State surveillance of social media user data may interfere with or enable interference with a wide range of human rights – for example, by discouraging freedom of expression, association or assembly, enabling discrimination and/or facilitating arbitrary detention, torture or extrajudicial killings.[166] To date, human rights courts and treaty bodies have primarily examined state surveillance through the prism of the right to privacy, as recognized in Article 17 of the ICCPR and a range of regional human rights treaties. In practice, since states typically have little difficulty identifying a legitimate aim for their surveillance practices – for example, the protection of national security – attention has generally been directed towards the tests of legality and necessity.[167] On these questions, the recent practice of the ECtHR has proven particularly instructive, addressing the tests of legality and necessity jointly by examining whether domestic law is accessible and foreseeable in application and contains adequate and effective safeguards and guarantees against abuse.[168] In the age of social media surveillance, however, three questions have risen in prominence concerning the application of these safeguards in practice.

First, as states have increasingly sought access to contextual communications data on social media platforms – including, for example, data identifying the sender, recipient, time, location and duration of a communication – the question of the extent to which such metadata is protected by the right to privacy has become increasingly salient. In practice, the aggregation of metadata may allow very precise conclusions to be drawn about an individual's behaviour, social relationships and identity that

[162] 'NSA Prism Program Taps in to User Data of Apple, Google and Others', *The Guardian* (7 June 2013).
[163] Rozenshtein, *supra* note 12, at 116–119.
[164] 'Facebook Says Government Demands for User Data Are at a Record High', *TechCrunch* (13 November 2019).
[165] Freedom House, *Freedom on the Net 2019: The Crisis of Social Media* (2019), at 12–20.
[166] *Ibid.*
[167] See, e.g., ECtHR, *Roman Zakharov v. Russia*, Appl. no. 47143/06, Judgment of 4 December 2015, para. 237.
[168] *Ibid.*, para. 236.

extend beyond even what may be conveyed by the content of a private communication.[169] With this in mind, it is both notable and welcome that there is now growing recognition within human rights courts that the acquisition of metadata is not necessarily less intrusive than the acquisition of the actual content of communications and should therefore be subject to adequate safeguards against abuse under HRL.[170]

Second, as states have increasingly established legislative frameworks that require providers of electronic communications services, including social media companies, to enable general and indiscriminate retention, analysis and/or transmission of metadata to their security and intelligence agencies, the question of the compatibility of such bulk surveillance activities with HRL has also grown in importance. To date, the two courts that have examined this question most extensively – the CJEU and the ECtHR – appear to be converging in their approach.[171]

According to the case law of the CJEU, the compatibility of bulk surveillance practices with the rights recognized in the Charter of Fundamental Rights of the European Union, including the rights to privacy and personal data protection, depends primarily on whether such practices are 'strictly necessary'.[172] Applying this test, the CJEU in *Tele2 Sverige* concluded that 'national legislation providing for the general and indiscriminate *retention* of all traffic and location data' was not strictly necessary to achieve the objective of *fighting organised crime and terrorism*.[173] More recently, the CJEU concluded that the general and indiscriminate *transmission* of traffic and location data to security and intelligence agencies for the purpose of *safeguarding national security* exceeds the limits of what is strictly necessary and therefore cannot be considered to be justified within a democratic society.[174] At the same time, in what appears to signal a softening of the restrictive stance elaborated in *Tele2 Sverige*, the CJEU adopted a differentiated approach whereby the permissibility of an instruction requiring providers of electronic communication services to *retain or analyse*, generally and indiscriminately, various types of communications data hinges on the type of communications data retained/analysed, the legitimate purpose relied upon and the adequacy of the safeguards in place. For instance, the general and indiscriminate retention of traffic and location data is not precluded provided that the state is confronted with 'a serious threat to national security that is shown to be genuine and present or foreseeable', the decision imposing such an instruction is subject to effective review by an independent

[169] The Right to Privacy in the Digital Age: Report of the UN High Commissioner for Human Rights, UN Doc. A/HRC/39/29 (3 August 2018), paras 6, 18.

[170] See, e.g., ECtHR, *Big Brother Watch and Others v. United Kingdom*, Appl. no. 58170/13, 62322/14 and 24960/15, Judgment of 13 September 2018, paras 356, 464; Case C-623/17, *Privacy International v. Secretary of State for Foreign and Commonwealth Affairs* (EU:C:2020:790), para. 71.

[171] See similarly Zalnieriute, 'The Future of Data Retention Regimes and National Security in the EU after the *Quadrature Du Net* and *Privacy International* Judgments', *ASIL Insights* (5 November 2020).

[172] Charter of Fundamental Rights of the European Union, OJ 2012 C 326/02. See, e.g., *Privacy International*, *supra* note 170, para. 67.

[173] Joined Cases C-293/12 and C-594/12, *Tele2 Sverige AB v. Post- och telestyrelsen and Secretary of State for the Home Department v. Tom Watson and Others* (EU:C:2016:970), para. 103 (emphasis added).

[174] *Privacy International*, *supra* note 170, para 81.

body whose decision is binding and the instruction is given 'only for a period that is limited in time to what is strictly necessary, but which may be extended if that threat persists'.[175] In this context, 'general and indiscriminate' data retention arises where there is a lack of 'objective criteria that establish a connection between the data to be retained and the objective pursued'.[176]

In a similar vein, the case law of the ECtHR also seems to be moving in a more permissive direction with respect to bulk surveillance practices. Although the ECtHR in *Szabó and Vissy* initially endorsed the CJEU's test of 'strict necessity',[177] more recently in *Big Brother Watch & Others*, the ECtHR appeared to afford states a wider margin of appreciation to adopt bulk surveillance measures.[178] According to the ECtHR, not only do bulk interception regimes to identify unknown threats to national security fall within states' margin of appreciation, but such regimes also constitute '*a valuable means* to achieve the legitimate aims pursued, particularly given the current threat level from both global terrorism and serious crime'.[179] To reach this conclusion, the ECtHR deferred to the findings of the UK's Independent Reviewer of Terrorism Legislation and a 2015 report by the European Commission for Democracy through Law (the Venice Commission).[180]

Reflecting on the ECtHR's deference in this context, it is regrettable that the Court did not engage further with the question of the necessity and proportionality of the bulk interception measures – for example, by clarifying the types of activity that constitute a threat to national security warranting extensive bulk surveillance measures. Indeed, as Daragh Murray and Pete Fussey argue, 'it is for the state *to demonstrate* the necessity for such powers, and to detail why traditional alternatives are inadequate'.[181] Moreover, the endorsement of bulk interception regimes also led the Court to relax some of the safeguards against abuse established in its case law in the context of *targeted* surveillance regimes.[182] Specifically, the ECtHR concluded that requiring objective evidence of reasonable suspicion in relation to the persons for whom data is being sought and the subsequent notification of the surveillance would be inconsistent with the Court's acknowledgement that the operation of a bulk regime in principle falls within a state's margin of appreciation because such requirements 'assume the existence of clearly defined surveillance targets, which is simply not the case in a bulk interception regime'.[183] The *Big Brother Watch* case is currently awaiting judgment before the Grand Chamber, and it remains to be seen whether the ECtHR's case law will seek to align its approach with the most recent findings elaborated by the CJEU.

[175] Joined Cases C-511/18, C-512/18 and C-520/18, *La Quadrature du Net et al. v. Order des Barreaux Francophones et Germanophone et al.* (EU:C:2020:791), Disposition.

[176] *Ibid.*, at para 133.

[177] ECtHR, *Szabó and Vissy v. Hungary*, Appl. no. 37138/14, Judgment of 12 January 2016, para. 73.

[178] Christakis and Bouslimani, 'National Security, Surveillance and Human Rights', *SSRN* (8 June 2020).

[179] *Big Brother Watch and Others*, *supra* note 170, paras 314, 386 (emphasis added).

[180] *Ibid.*, paras 176, 211, 384–385.

[181] Murray and Fussey, 'Bulk Surveillance in the Digital Age: Rethinking the Human Rights Law Approach to Bulk Monitoring of Communications Data', 52 *Israel Law Review* (2019) 31, at 58 (emphasis added).

[182] See similarly Christakis and Bouslimani, *supra* note 178.

[183] *Big Brother Watch and Others*, *supra* note 170, para. 317.

Finally, given that cyber surveillance campaigns are often conducted beyond a state's territorial borders, questions have also arisen concerning the extent to which HRL applies extraterritorially. Article 2(1) of the ICCPR, for example, provides that each state party 'undertakes to respect and to ensure to all individuals *within its territory and subject to its jurisdiction* the rights recognised in the present Covenant'.[184] According to the predominant view, a state must respect and ensure the rights of individuals physically located beyond its territorial borders when the state exercises 'power or effective control' either over the territory on which the individual is located (the *spatial* model of jurisdiction) or over the individual (the *personal* model of jurisdiction).[185]

The effective control test was initially developed with respect to situations where a state had *physical* control over a territory or an individual. The virtual nature of cyber surveillance operations, however, raises the question of whether *virtual* control suffices to satisfy the test. On this question, opinion is divided. While some authorities suggest that physical control is required,[186] there is also emerging support for a more flexible reading of the effective control test tailored to the technological advances of the social media age. The UN HRC, for example, has concluded that states should adopt measures to ensure that any interference with the right to privacy complies with the tripartite tests of legality, legitimacy and necessity 'regardless of the nationality or location of the individuals whose communications are under direct surveillance'.[187] The Office of the UN High Commissioner for Human Rights has also concluded that cyber surveillance practices may engage a state's human rights obligations 'if that surveillance involves the State's exercise of power or effective control *in relation to digital communications infrastructure, wherever found*, for example, through direct tapping or penetration of that infrastructure'.[188] Although these findings remain at a nascent stage of development, there are also indications of similar flexibility at the regional level. The Inter-American Court of Human Rights, for example, recently concluded – in the context of an advisory opinion concerning the environmental obligations of states – that jurisdiction arises pursuant to Article 1(1) of the American Convention on Human Rights 'when the State of origin exercises effective control *over the activities carried out that caused the harm and consequent violation of human rights*'.[189] Similarly, in *Ilascu*

[184] ICCPR, *supra* note 52, Art. 2(1) (emphasis added).

[185] UN HRC, General Comment no. 31: The Nature of the General Legal Obligations Imposed on States Parties to the Covenant, UN Doc. CCPR/C/21/Rev1/Add.13 (29 March 2004), para. 10. A minority position, put forward by, *inter alia*, the United States, submits that human rights do not apply extraterritorially. See UN HRC, 'Summary Record of the 1405th Meeting', UN Doc. CCPR/C/SR/1405, 24 April 1995, at para. 20.

[186] M.N. Schmitt (ed.), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2017), at 185 (in which a majority of the experts argue that physical control is required).

[187] UN HRC, Concluding Observations on the Fourth Periodic Report of the United States of America, UN Doc. CCPR/C/USA/CO/4 (23 April 2014), para. 22.

[188] Report of the Office of the UN High Commissioner for Human Rights, UN Doc. A/HRC/27/37 (30 June 2014), para. 34 (emphasis added).

[189] IACtHR, *Environment and Human Rights*, Advisory Opinion OC-23/17 (15 November 2017), para.104(h) (author's translation; emphasis added); see also UN HRC, General Comment no. 36 on Article 6 of the International Covenant on Civil and Political Rights, on the Right to Life, UN Doc. CCPR/C/GC/36 (30 October 2018), para. 22. American Convention on Human Rights 1969, 1144 UNTS 123.

and Others v. Moldova and Russia, the ECtHR confirmed that a state's responsibility may be engaged 'on account of acts which have sufficiently proximate repercussions on rights guaranteed by the Convention, even if those repercussions occur outside its jurisdiction'.[190] While the precise boundaries of these tests remain to be clarified through case law, they signal a judicial openness to expanding the extraterritorial application of HRL by focusing on whether the state has effective control over the enjoyment of the rights of individuals – an approach that would enable the applicability of HRL to most forms of cross-border cyber surveillance operations.[191]

Questions concerning whether and which safeguards are adequate under HRL to legitimate different types of bulk surveillance practices, as well as uncertainties that persist concerning the extraterritorial scope of HRL, are undoubtedly significant in the age of social media surveillance. At the same time, it is important to remember that these questions sit firmly within a marketized perspective of HRL that leaves relatively untouched the underlying social media ecosystem that enables the accumulation and centralization of so much data in the first place. Bearing this in mind, it is notable that, while social media companies have proven willing and sometimes enthusiastic to recognize and add friction to the human rights threats posed by state surveillance of their platforms, they have been far more reluctant to acknowledge their role as surveillance principals whose corporate practices themselves negatively interfere with individual rights.[192] This reticence is all the more concerning given the increasingly ubiquitous and unaccountable nature of platform surveillance. If HRL is to avoid crowding out and diverting attention from the underlying architecture of platform surveillance, a more structural perspective is required that addresses concerns associated with the data extractive business models of social media companies.

B *Surveillance Principals*

Under HRL, the regulation of social media companies as surveillance principals has primarily been addressed through the prism of informational privacy, which forms the foundation of data protection law.[193] Although data protection has been recognized in some constitutional systems as a distinct right,[194] under HRL the protection of personal data has traditionally been regarded as a component of the right to privacy.[195] In fact, it is now well established that states are under a positive obligation to regulate the processing of personal data in order to ensure the enjoyment of the right to privacy.[196]

[190] ECtHR, *Ilascu and Others v. Moldova and Russia*, Appl. no. 48787/99, Judgment of 8 July 2004, para. 317.

[191] B. Çali, 'Has "Control over Rights Doctrine" for Extra-Territorial Jurisdiction Come of Age? Karlsuhe, Too, Has Spoken, Now It's Strasbourg's Turn', *EJIL:Talk!* (21 July 2020).

[192] Jørgensen, 'Rights Talk: In the Kingdom of Online Giants', in R.F. Jørgensen (ed.), *Human Rights in the Age of Platforms* (2019) 163.

[193] Amnesty International, *supra* note 159, at 19–22.

[194] See, in particular, Charter of Fundamental Rights, *supra* note 172, Art. 8(1).

[195] UN HRC, General Comment no. 16: Article 17 (The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation), Doc. HRI/GEN/1/Rev.9 (Vol. I) (8 April 1988), para. 10; ECtHR, *I. v. Finland*, Appl. no. 20511/03, Judgment of 17 July 2008, para. 38.

[196] General Comment no. 16, *supra* note 195, para. 10; *I. v. Finland*, *supra* note 195, paras 35–49; ECtHR, *K.U. v. Finland*, Appl. no. 2872/02, Judgment of 2 December 2008, paras 40–50; see also Zingales, 'A Stronger Claim to Data Protection during Pandemics? Leveraging the American Convention of Human Rights against Government Inaction: A Brazilian Case Study', *SSRN* (27 September 2020).

As the Office of the UN High Commissioner for Human Rights recently explained, 'the track record to date implying mass, recurrent misuse of personal information by some business enterprises confirms that legislative measures are necessary for achieving an adequate level of privacy protection'.[197] It remains an open question, however, what an 'adequate level' of protection means as a matter of HRL in the age of platform surveillance.

In practice, data protection regimes around the world predominantly adhere to marketized conceptions of HRL, premised on 'neoliberal models of agency' and 'a marketplace model of enlightenment' that place substantial faith in the capacity of individuals to seek out and interpret information about the online environment, make informed choices about it and exercise rights in relation to it.[198] According to these 'privacy self-management' approaches to data protection, most forms of data processing are permissible provided individuals are notified and provide consent.[199] In the contemporary social media era, notice and consent models of data protection are problematic to the extent that they neglect structural asymmetries of power between users and platforms, render personal data a mere commodity that can be traded for access to online services, inadequately account for the effects of individual consent on third parties and occlude consideration of whether it is appropriate for individual consent to operate as a form of legitimation of harms associated with the platform economy.[200] As Lillian Edwards and Michael Veale observe, '[c]onsent as an online institution in fact arguably no longer provides any semblance of informational self-determination but merely legitimises the extraction of personal data from unwitting data subjects'.[201]

Looking to the future, it remains to be seen whether more recent approaches to data protection, most notably the EU's General Data Protection Regulation (GDPR), will mark a shift away from the model of notice and consent.[202] Structurally, the GDPR establishes a multi-layered system of governance underpinned by a set of personal data protection principles,[203] which are given more detailed expression in the form of data subject rights,[204] data controller and processor obligations and responsibilities[205] and

[197] Right to Privacy in the Digital Age, *supra* note 169, para. 27.

[198] Annany and Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability', 20 *New Media and Society* (2018) 973, at 979.

[199] Solove, 'Privacy Self-Management and the Consent Dilemma', 126 *Harvard Law Review* (2013) 1880, at 1883–1893.

[200] See generally Bietti, 'Consent as a Free Pass: Platform Power and the Limits of the Informational Turn', 40 *Pace Law Review* (2020) 307.

[201] Edwards and Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For', 16 *Duke Law and Technology Review* (2017) 18, at 66; see also Belli *et al.*, 'Selling Your Soul While Negotiating the Conditions: From Notice and Consent to Data Control by Design', 7 *Health and Technology* (2017) 453.

[202] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation (GDPR)), OJ 2016 L 119/1.

[203] *Ibid.*, Art. 5.

[204] *Ibid.*, Arts 12–23.

[205] A data controller is a natural or legal person that 'determines the purposes and means of the processing of personal data', while a data processor is a natural or legal person which 'processes personal data on behalf of the controller'. *Ibid.*, Art. 4(7)–(8).

public-private co-regulatory partnerships that aim to systemically improve personal data processing practices.[206] Yet, while consent is not the only lawful basis for data processing, the extent to which the GDPR will move beyond the privacy self-management approach to data protection will ultimately depend on how its provisions are interpreted and enforced in practice.

In terms of *interpretation*, the GDPR contains a number of open-textured provisions and thresholds that require clarification in the context of concrete cases. According to Chris Hoofnagle, Bart van der Sloot and Frederik Borgesius, for example, we are currently at the very beginning of 'an extended tussle between authorities and large companies such as Google and Facebook that involves positioning, anchoring, and other gamesmanship intended to blunt the GDPR's effects'.[207] Initial indications, however, suggest that data protection authorities are tending to focus narrowly on specifying the requirements for companies to rely on informed consent as a lawful basis of data processing rather than questioning whether consent is a suitable basis for legitimizing data processing given the structure of the contemporary online environment.[208]

France's data protection authority, the Commission Nationale de l'Informatique et des Libertés (CNIL), for example, recently fined Google 50 million euros for failing to adhere to the requirements for valid consent under the GDPR.[209] The CNIL concluded that Google had breached its transparency and information obligations because essential data processing information was excessively scattered across several documents and webpages, whilst also being insufficiently clear and understandable. These breaches meant that user consent for the processing of personalized advertising was also not properly informed, while Google's reliance on pre-ticked boxes to obtain consent to process data for behavioural targeting rendered consent inadequately specific and unambiguous. These findings are revealing both for what they include and exclude. In terms of inclusion, the CNIL's findings focus narrowly on how Google could improve the presentation of information in order to rely on consent as a lawful basis of data processing on the implicit assumption that user consent may be perfected in this context.[210] At the same time, the CNIL neglected to consider whether Google's market-dominant position amounts to 'a clear imbalance' of power between the data subject and the controller such that consent could not be characterized as freely given.[211] The result, as Elettra Bietti explains, is that the CNIL's decision seems to 'assume that the individual can and should be the ultimate decision-maker regarding opaque commercial data practice … [whilst] neglecting the power asymmetries and information externalities that make individual-centric decision-making objectionable'.[212]

[206] *Ibid.*, Arts 25, 35–43; see also Edwards and Veale, *supra* note 201, at 74–80.

[207] Hoofnagle *et al.*, 'The European Union General Data Protection Regulation: What It Is and What It Means', 28 *Information and Communications Technology* (2019) 65, at 69.

[208] Bietti, *supra* note 200, at 339.

[209] Commission Nationale de l'Informatique et des Libertés, *Délibération de la formation restreinte n° SAN – 2019-001 du 21 janvier 2019 prononçant une sanction pécuniaire à l'encontre de la société GOOGLE LLC*, Doc. SAN-2019-001 (21 January 2019).

[210] Bietti, *supra* note 200, at 342.

[211] GDPR, *supra* note 202, Recital 43.

[212] Bietti, *supra* note 200, at 339.

In terms of *enforcement*, the effectiveness of the GDPR will depend on a range of practical factors, including whether the opportunities provided by the GDPR for member state derogations, exceptions and restrictions will enable social media companies to minimize their regulatory burden through arbitrage,[213] whether the GDPR's minimal involvement of third parties within its co-regulatory processes, such as codes of conduct, impact assessments and certification mechanisms, will render them susceptible to regulatory capture[214] and whether data protection authorities will be sufficiently well funded and resourced to enforce the GDPR.[215]

Importantly, it is entirely possible that the GDPR could be interpreted and enforced in alignment with a more structural conception of HRL, one where consent plays only a marginal role in legitimating data processing practices within the contemporary social media ecosystem. For example, greater emphasis might be placed on the GDPR's provisions concerning data protection by design and by default, the principles of data integrity, confidentiality and minimization and data protection impact assessments, each of which could be utilized to structurally improve the design and oversight of data processing in the platform economy.[216] At the same time, data protection authorities could reject explicit consent as a lawful ground of data processing for online behavioural microtargeting,[217] a move that could potentially trigger a structural shift away from an advertising system that relies on processing enormous amounts of personal data towards, for example, a model that targets advertising contextually based on location and real-time interests.[218]

In steering data protection regimes such as the GDPR in this direction, the wider web of human rights courts, treaty bodies and experts has a potentially significant role to play in specifying what constitutes an adequate level of data protection under HRL in the context of platform surveillance. For instance, rather than limiting its discussions of privacy to narrower issues of state surveillance and data retention legislation,[219] the UN HRC could devote greater space in its concluding observations to questions concerning the interpretation and enforcement of data protection regimes in light of the structure of the contemporary platform economy. Only by adhering to a more structural conception of HRL can data protection regimes such as the GDPR guard against legitimating merely cosmetic changes to the current social media

[213] Access Now, *One Year under the EU GDPR: An Implementation Progress Report* (2019), at 3.

[214] Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability', 92 *Southern California Law Review* (2019) 1529, at 1608–1610.

[215] Edwards and Veale, *supra* note 201, at 75.

[216] See similarly Bietti, *supra* note 200, at 394.

[217] See, however, Information Commissioner's Office (ICO), *Update Report into AdTech and Real Time Bidding* (20 June 2019), at 16 ('[m]arket participants must therefore modify existing consent mechanisms to collect explicit consent, or they should not process this data at all').

[218] Lomas, 'Behavioural Advertising Is Out of Control, Warns UK Watchdog', *TechCrunch* (20 June 2019); see also Maréchal, MacKinnon and Dheere, *supra* note 39, at 34–35; and Balkin, *supra* note 157, at 27–29.

[219] See, e.g., UN HRC, Concluding Observations on the Fourth Periodic Report of Estonia, UN Doc. CCPR/C/EST/CO/4 (18 April 2019), at paras 29–30.

ecosystem at the expense of addressing more systemic concerns associated with the data extractive business models of today's leading social media companies.

4 Conclusion

The kinetic growth of the online platform economy over the course of the past 15 years has been infused with the values of the neoliberal era in which it emerged. The result has been the privatization, commodification and datafication of the digital public sphere. Set in this context, this article has examined the relationship between HRL and the contemporary social media ecosystem. Recognizing that HRL is a vocabulary of governance with the potential to both restrain and legitimate particular relations of power within the platform economy, this article has revealed not only the contestability of HRL but also the inadequacies of adopting a marketized conception of HRL to address the accountability deficits associated with social media platforms.

By prioritizing the negative obligations of states to refrain from unjustifiable interferences with human rights, a marketized understanding of HRL is ill-equipped to confront the privatized moderation and surveillance practices of social media companies or the increasingly informalized forms of pressure exerted by states to influence such practices. In addition, by relying on a narrow abstract individualism, a marketized conception of HRL also fails to adequately account for the background context and systemic effects of state and platform practices on freedom of expression and privacy across the social media ecosystem as a whole. This is particularly problematic where, for example, intermediary liability laws that incorporate pro-active monitoring obligations and encourage the use of pre-emptive filtering technologies to guard against difficult to adjudicate categories of unlawful content such as discriminatory hate speech are legitimized without accounting for the attendant systemic risks of over-removing lawful content. Equally troubling are data protection regimes that foreground individual consent as a lawful basis for data processing without accounting for systemic imbalances of power between platforms and users or reflecting on whether consent should be permissible where it functions primarily as a monetization model for online services and a mechanism for legitimating online harms associated with platform surveillance.

This article has suggested that shifting to a more structural conception of HRL would begin to address several of these concerns. In the content moderation context, for example, a structural conception of HRL would insist on a more holistic and evidence-based approach to the design of intermediary liability laws that strives to account for the systemic effects of such frameworks on online expression. A structural approach would also place greater emphasis on the positive obligation to protect freedom of expression as a basis for requiring states to ensure that robust mechanisms of transparency, due process, accountability and oversight are embedded in platform moderation systems as well as any public-private or cross-platform collaborative initiatives that are relied upon to influence content governance in this context. The positive obligation to protect freedom of expression also provides a basis for directing state

attention towards the challenge of identifying regulatory avenues for ensuring a more pluralized social media landscape. In terms of data surveillance, a structural conception of HRL would require states to establish data protection regimes that account for the asymmetries of power that exist between social media platforms and individual users – for example, by placing greater emphasis on principles such as data protection by design that seek to systemically challenge the data extractive business models of social media companies and improve the design and accountability of data processing in the contemporary online environment.

Of course, there is no guarantee that a structural conception of HRL will be adopted in practice. In this regard, it is important to emphasize that the sites for advancing a structural HRL agenda are not confined to courts but also encompass a broader array of political and legislative arenas. As Amy Kapczynski explains, '[a] revised human rights would not ignore courts – for there the battle can be lost, if never won – but must also be attentive to the need to build a broader politics, and structures of political accountability that are needed to achieve a more ambitious vision of justice at a global scale'.[220] Nor should it be thought that a structural conception of HRL offers a uniform set of answers for the accountability challenges of the online platform ecosystem. Rather, the principal value of a structural conception of HRL resides in more modestly offering a way of thinking about the relationship between HRL and platform governance that is more attentive to the systemic dimensions and political economy of the online platform ecosystem.

Finally, it is also important to recognize that a structural conception of HRL is not a panacea and constitutes only one limited regulatory vocabulary amongst many through which to address the democratic concerns associated with social media platforms. Other regulatory terrains include the broader normative vocabulary of human duties and responsibilities – encompassing not only the corporate responsibility to respect human rights elaborated in the UN Guiding Principles on Business and Human Rights, but also the complementary duties and responsibilities of the wider set of actors that participate within the social media ecosystem including media organizations, civil society groups and individuals – as well as the domains of social, political and economic policy.[221] In this regard, it is important to emphasize that many of the democratic concerns associated with social media platforms rely on exploiting societal fault lines rooted in structural dynamics that have been undermining democratic principles in societies around the world for decades. In the USA, for example, democratic principles have been undermined by longer-term dynamics of political economy, including decades of media deregulation, structural inequalities in the electoral system, policy disasters such as the Iraq War and the adoption of policies that have exacerbated economic inequality.[222] Going forward, addressing these structural concerns will require a perspective that extends beyond any single regulatory paradigm or any particular technological medium.

[220] Kapczynski, *supra* note 14, at 95.

[221] See, e.g., Helberger *et al.*, 'Governing Online Platforms: From Contested to Cooperative Responsibility', 34 *The Information Society* (2018) 1; Land, 'Speech Duties', 112 *AJIL Unbound* (2018) 329, at 329.

[222] Balkin, 'Constitutional Crisis and Constitutional Rot', 77 *Maryland Law Review* (2017) 147, at 157ff.